OFFICE OF PEOPLE ANALYTICS

CAT-ASVAB Pools 11-15 Equating

Matt Reeder

Human Resources Research Organization (HumRRO)

Briefing presented to the DACMPT
August 16, 2023

Objectives

- Pool development process overview
- CAT-ASVAB Pools 11–15 background
- Equating objective
- The reference scale
- CAT-ASVAB equating study
- ASVAB equating results
- ASVAB equating evaluation analyses
- Questions/Discussion

CAT-ASVAB Pool Development Process Overview

Data Collection & Processing
- Administer tryout items to applicants
- Process/clean data

Calibration & Scaling
- Calibrate 3PL parameters
- Rescale to operational

Item Screening
- Psychometric quality review
- Differential item functioning

Item Enemy Identification
- MC & MK: Rating
- Other tests: NLP-based process

Consolidate Analysis Results
- Eligible items
- Enemy relationships
- Content taxonomy

CAT Pool Assembly
- Simulation-based optimization algorithm
- Maximize score information/parallel pools

Additional CAT Parameters
- Exposure control
- Penalty parameters

Equating
- Conducted in three phases
- Equating evaluation

OPA OFFICE OF PEOPLE ANALYTICS

CAT-ASVAB Pools 11–15 Background

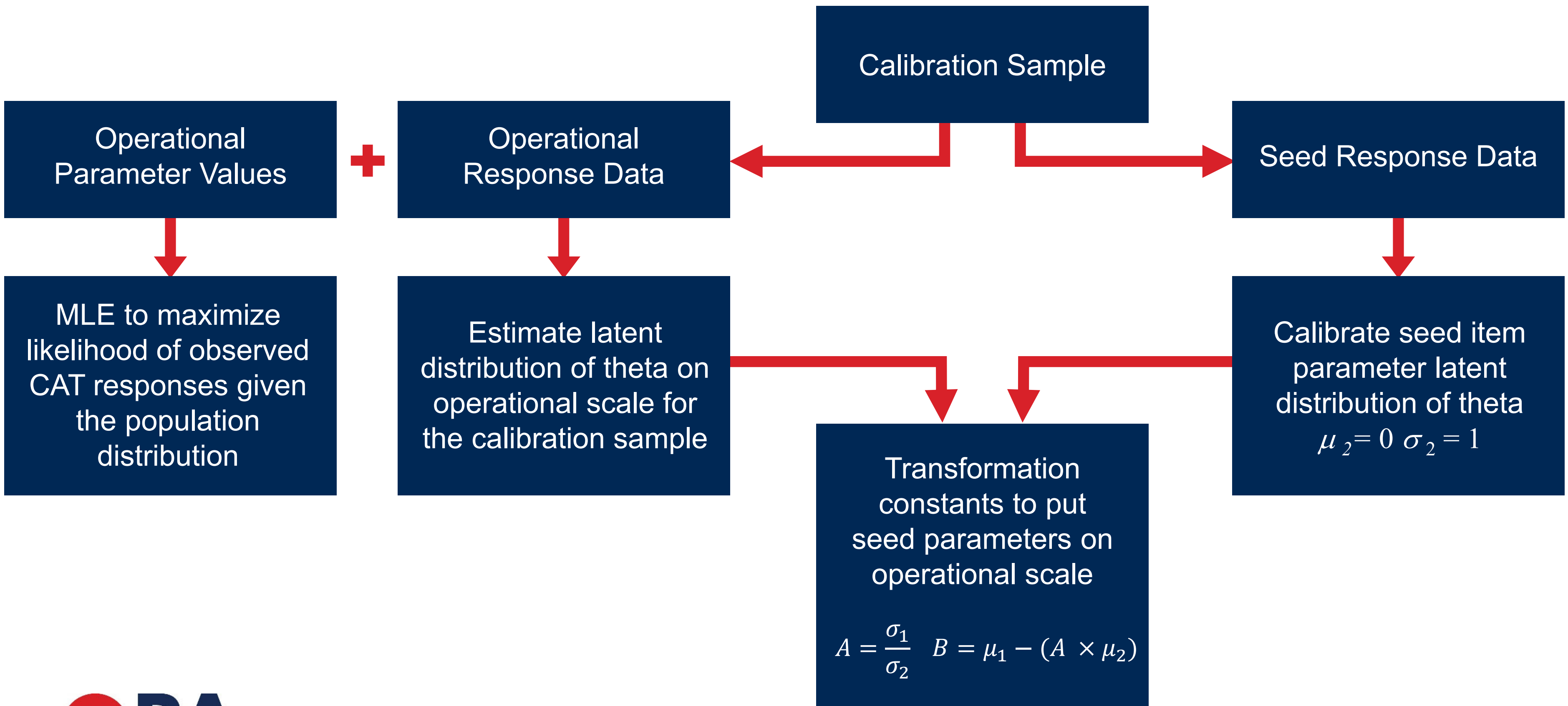
- CAT-ASVAB 11–15 are new CAT-ASVAB pools developed from tryout/seeded items that have passed content, psychometric, and fairness/sensitivity evaluations
- Items are assigned to pools through an optimization algorithm designed to maximize conditional precision levels of each pool and to constrain conditional precision levels to be comparable across pools
- Each CAT-ASVAB pool is unique (i.e., no common items)
- Item replacement is performed at the pool level
  - Operational CAT-ASVAB pools are typically repurposed when new pools are implemented
  - New and prior pools do not share any items
- Item parameters of items included in pools 11–15 have been rescaled to the operational
  CAT-ASVAB scale (i.e., CAT-ASVAB pools 5–9) prior to the equating study
- Standard scores (SS) are generated via a linear transformation, where the mean and standard deviation of the SS of new pools are matched to those of the reference pool
- Linear transformation constants to transform $\hat{\theta}$ to SS are estimated during the equating study

CAT-ASVAB Pools 11–15 Background: Item Parameter Rescaling

- Calibrate **seed** parameter values using seed response data. Latent distribution of theta is fixed to BILOG defaults (0,1)
- Use **operational** responses from calibration sample + operational parameter values to estimate latent distribution of theta on the operational scale for the calibration sample
- Compute transformation constants to put seed parameters on the operational scale

OPA
OFFICE OF PEOPLE ANALYTICS

ITEM PARAMETER RESCALING PROCESS



OPA
OFFICE OF PEOPLE ANALYTICS

Equating Objective: Standard Score Equating

Equipercentile Objective:

- ASVAB forms/pools have historically been equated to a reference form/pool using equipercentile methods to produce **equivalent composite distributions** across alternate forms/pools
  - When ASVAB transitioned to Item Response Theory (IRT) scoring, new CAT-ASVAB pools continued to be equated to a reference pool using a linear method that matches the mean and standard deviation of standard scores to a reference pool
- The current equating approach relies more heavily on the invariance property of IRT **and** aims to create equal distributions of scores across alternate pools
- Large volume of applicants qualify on CAT-ASVAB, and small differences between (unequated) pools can potentially have a large impact on the number of qualified applicants
- Composite cut scores should achieve the same selection ratio across pools
- Effectiveness of score equating is evaluated on the extent to which the pools can be used interchangeably to qualify the same proportion of applicants using selection composites

The Reference Scale: CAT-ASVAB Pool 4

- The ASVAB score scale allows policy makers to compare current applicant aptitude with past applicants, and to set target qualifications accordingly (Segall, 2004)
- The current ASVAB score scale was developed from a nationally representative sample collected during the 1997 Profile of American Youth (PAY97) study
  - See DAC briefing (Baumer, 2022) for re-norming needs assessment details
- Changes to ASVAB, like introducing new CAT pools, must be introduced in a deliberate, carefully planned manner to ensure the continuity of the interpretation of ASVAB scores
- Any given composite cut score should have the same meaning . . .
  - irrespective of which pool is administered
  - as it did when standards were originally set
- CAT-ASVAB pool 4 was included in the PAY97 norming study (Moore, Pedlow, & Wolter, 1999)
- CAT-ASVAB pool 4 has subsequently been administered for special purposes only and serves to define the reference scale for future equating studies

CAT-ASVAB EQUATING STUDY

- Rigorous equating procedures were developed by DTAC to equate pools 5–9 and put then onto the CAT-ASVAB scale
  - Used this as template for equating pool 10
  - Also a template for new pools 11–15
- Conducted at the subtest level
- Linear equating methods used to derive constants to transform IRT-based theta scores ($\hat{\theta}$) on pools 11–15 to scale of the reference pool 4 in a phased approach
- Random groups design
  - Each applicant is assigned to a single pool with 1/7 assignment probability
    - The reference pool (4), administered only during equating studies
    - An operational pool (5)
    - A new pool (11–15)
- Evaluate differences in qualification composite cumulative distribution functions (CDFs) between reference pool 4 and new pools

CAT-ASVAB Equating Study

- Equating is implemented in three phases of operational administration of new pools to military applicants
  - Each phase includes progressively larger sample size
  - Phase sample sizes are cumulative such that they include all individuals from the previous phase
  - Intent of phased design is to maximize accuracy of reported operational scores
- **Phase 1:** Provisional equating based on IRT invariance
- **Phase 2:** Use data from phase 1 to update combined (across pool) transformation constants
- **Phase 3:** Use data from phase 2 to update/refine combined (across pool) transformation constants
- **Final:** Use data from phase 3 to estimate final separate (pool-specific) transformation constants to be applied to applicants testing post initial operational test and evaluation (IOT&E)

ASVAB Equating: Data Collection N's

Pool	Description	Assignment Probability	Phase 1 Target | Actual	Phase 2 Target | Actual	Phase 3 Target | Actual
4	Reference	**1/7**	500 | 534	1,500 | 1,633	10,000 | 10,031
5	Operational	**1/7**	500 | 499	1,500 | 1,596	10,000 | 10,078
11	New	**1/7**	500 | 527	1,500 | 1,626	10,000 | 10,295
12	New	**1/7**	500 | 495	1,500 | 1,561	10,000 | 10,218
13	New	**1/7**	500 | 548	1,500 | 1,659	10,000 | 10,228
14	New	**1/7**	500 | 505	1,500 | 1,539	10,000 | 10,119
15	New	**1/7**	500 | 511	1,500 | 1,666	10,000 | 10,121

NOTE: N's across phases are cumulative. For example, the 1,633 examinees listed for pool 4 in phase 2 includes the 534 examinees from phase 1.

OPA
OFFICE OF PEOPLE ANALYTICS

ASVAB Equating: Phase 3 Analyses

- Random group equivalence
- Equating transformation constant estimation
- Pool subtest intercorrelation equivalence analysis
- Composite distribution equivalence
- Subgroup performance across pools
- Operational pool comparison
- Provisional equating transformation accuracy

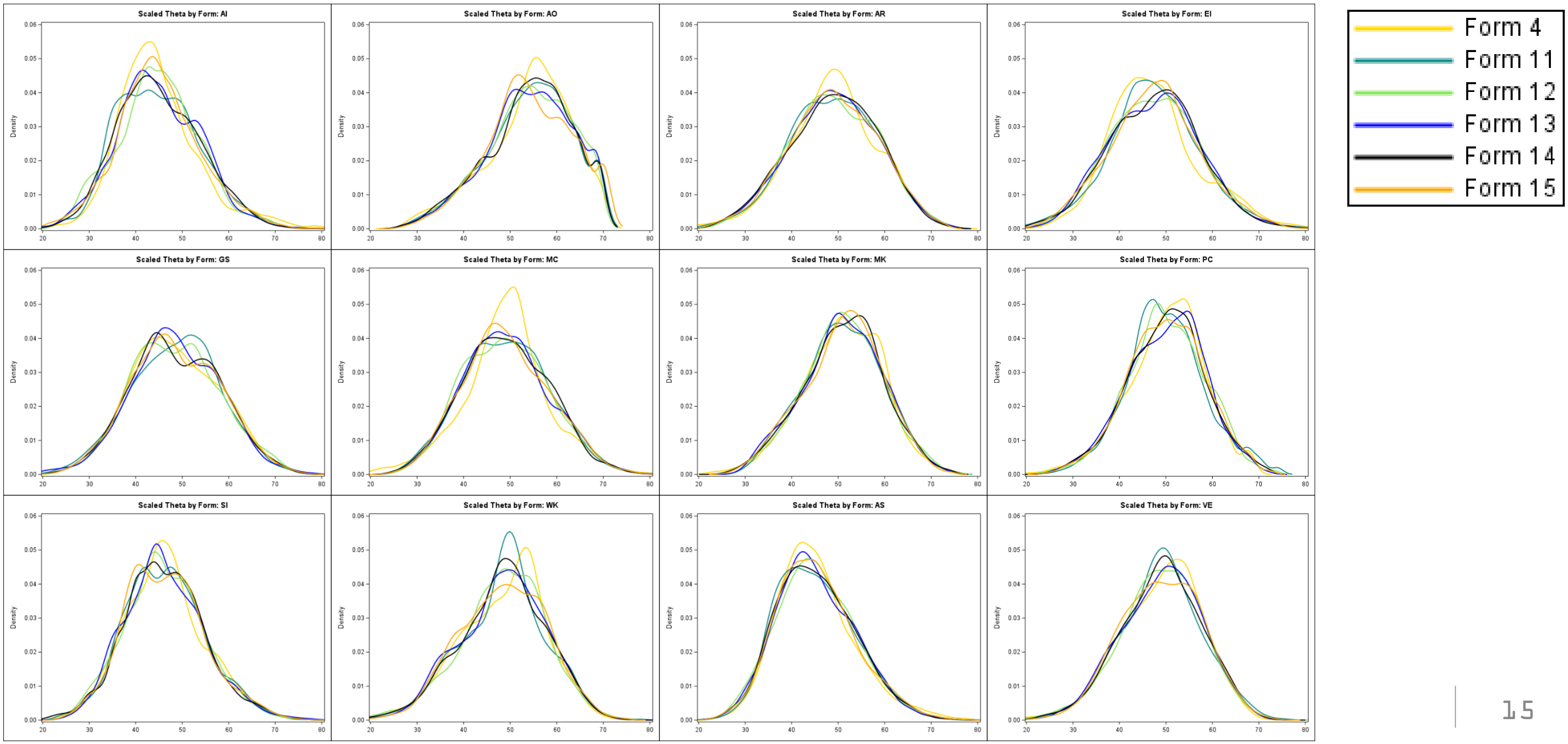
Random Group Equivalence

- Does assignment procedure produce equivalent groups with respect to key demographic variables?
- Compare distributions of key demographic variables across assignment to pools 4 and 11–15
  - Gender: χ^2 (5, *N* = 60,810) = 12.55, *p* = 0.03
  - Race: χ^2 (10, *N* = 57,489) = 2.44, *p* = 0.99
  - Ethnicity: χ^2 (5, *N* = 60,045) = 1.13, *p* = 0.95
- Expect groups assigned to different pools to be randomly equivalent
- Results suggest groups are randomly equivalent
  - α =.01 per extremely large sample size
  - Effect sizes (phi, V) associated with gender analysis are negligible (.01)

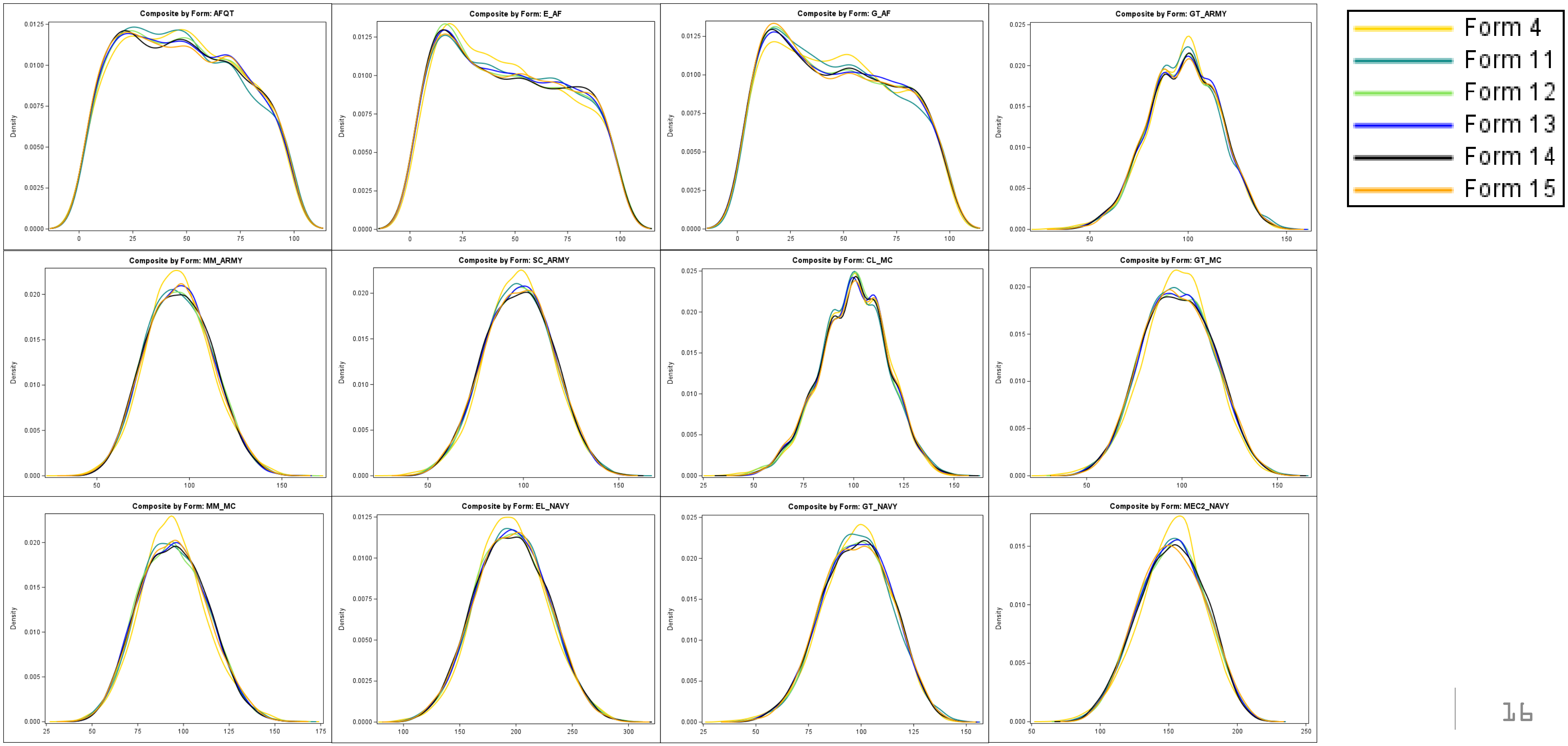
EQUATING TRANSFORMATION CONSTANT ESTIMATION

- Is linear transformation adequate?
- Do subtest distributions have similar shapes?
- Evidence of some systematic difference in shapes of subtest distributions
  - Not problematic for ASVAB
  - Qualification decisions are based on composite scores
  - Composites are likely to be more normal-like, and evidence suggests this is indeed the case

ASVAB Equating: Subtest Distributions



ASVAB EQUATING: COMPOSITE DISTRIBUTIONS (SELECT EXAMPLES)

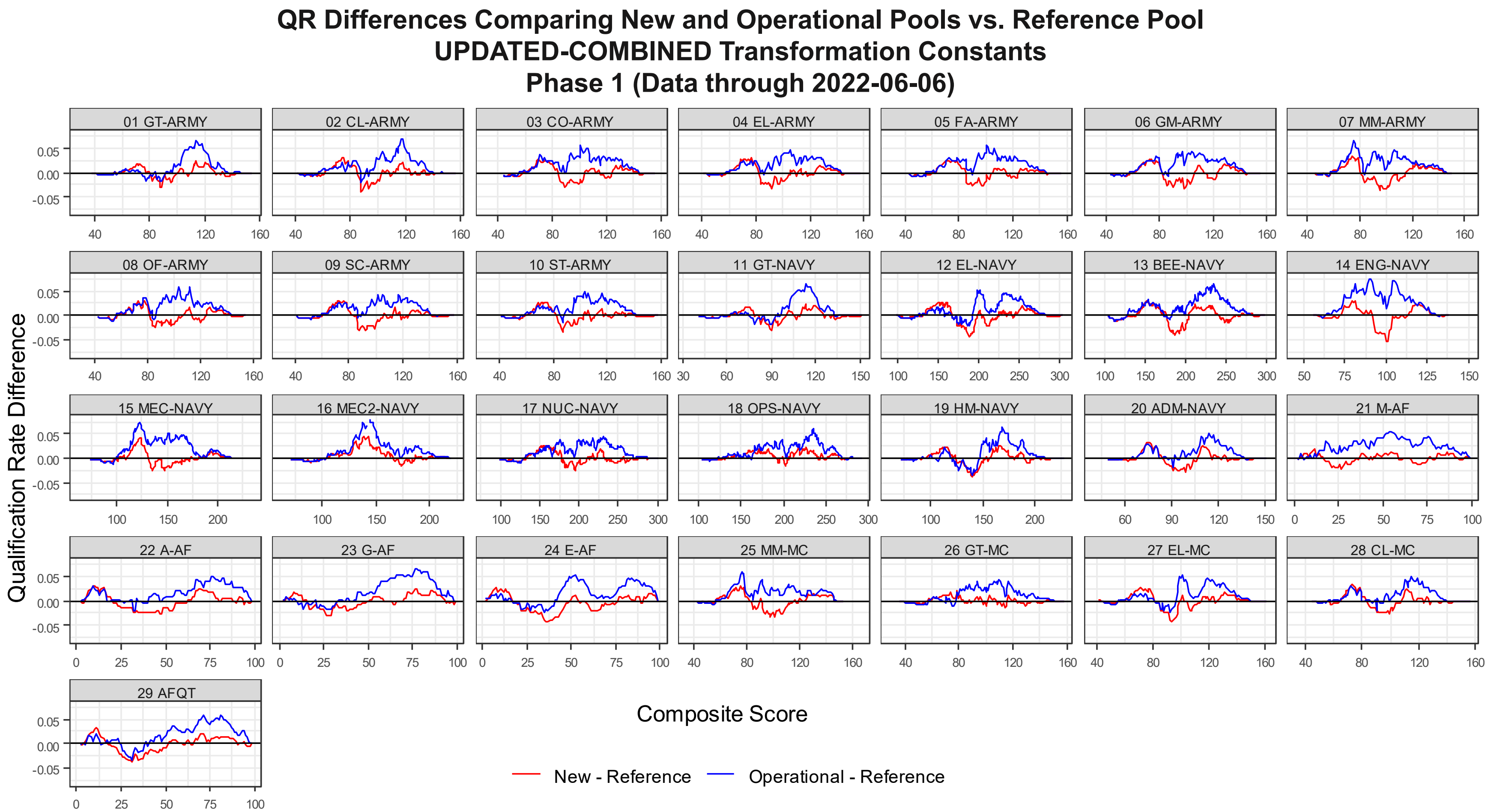


Pool Composite Equivalence Analysis

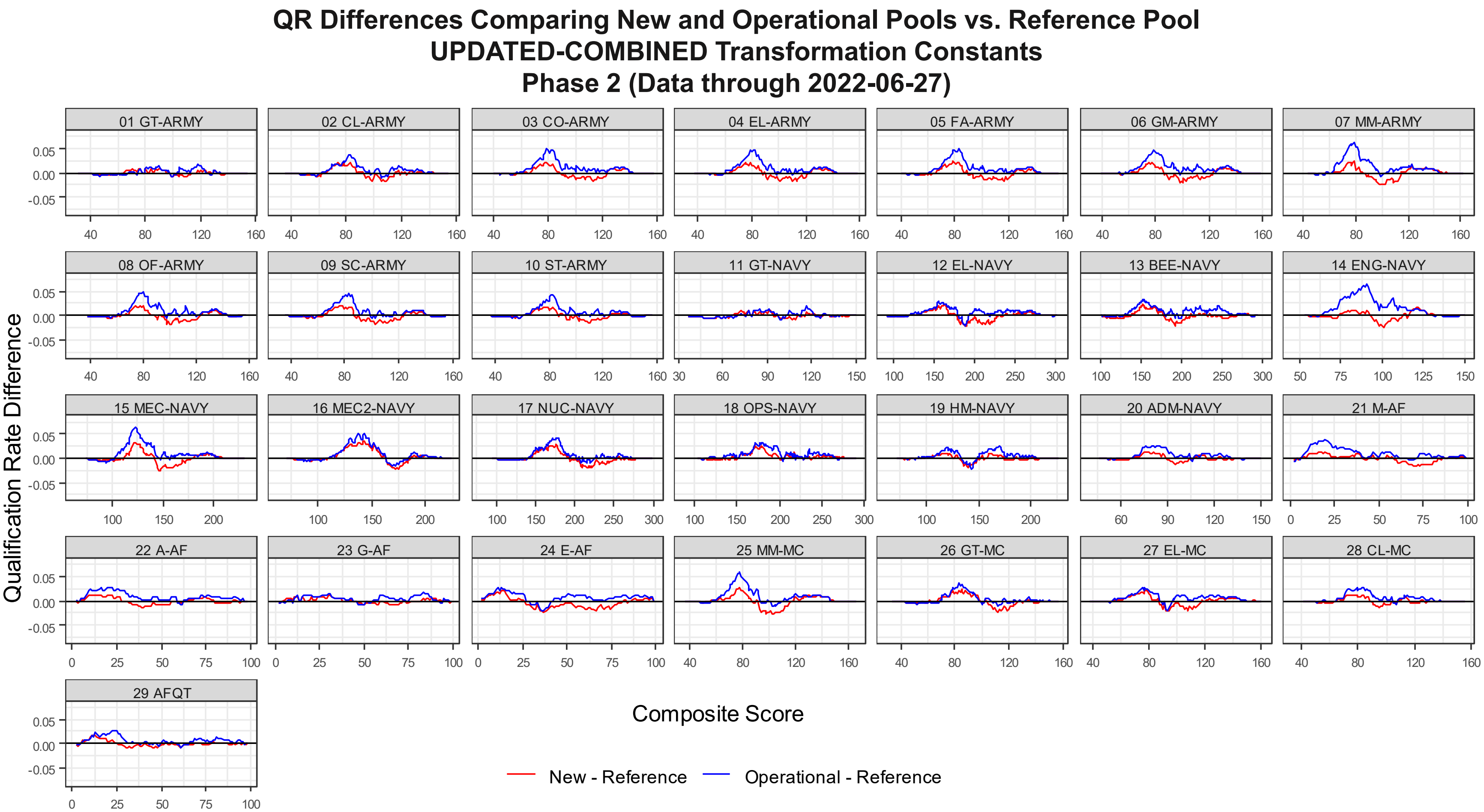
- Composites can have different variances if the pools display different patterns of subtest correlations
- Evaluate differences in cumulative distribution functions
  - Kolmogorov-Smirnov (K-S) test
  - Cumulative Distribution Function (CDF) for reference group minus CDF for new pool group
- Most composites displayed similar distributions across new and reference pools
  - Differences between new and reference pool CDFs decrease in each phase
- Five composites consistently (i.e., on average) displayed statistically significant differences between new and reference pool CDFs; however, these differences are
  - Comparable to what was observed during CAT-ASVAB Pool 5–9 equating
  - Relatively small
  - Within a tolerable range
- $max_{x\ 11:15} \left|CDF_{new}(x) - CDF_{ref}(x)\right| \leq 0.035$
- $max_{x\ 5:9} \left|CDF_{new}(x) - CDF_{ref}(x)\right| \leq 0.036$

OPA
OFFICE OF PEOPLE ANALYTICS

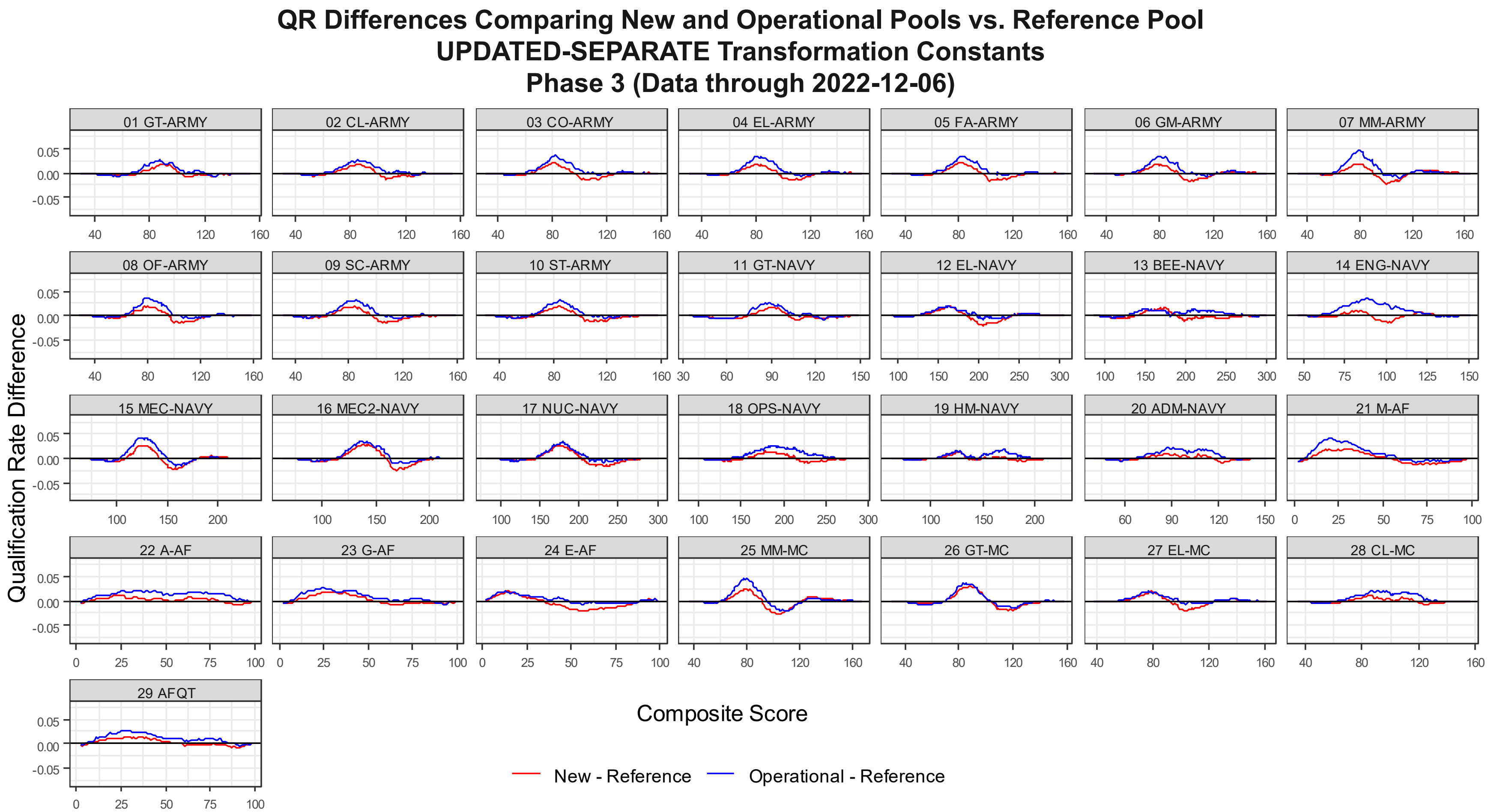
COMPOSITE DISTRIBUTION EQUIVALENCE



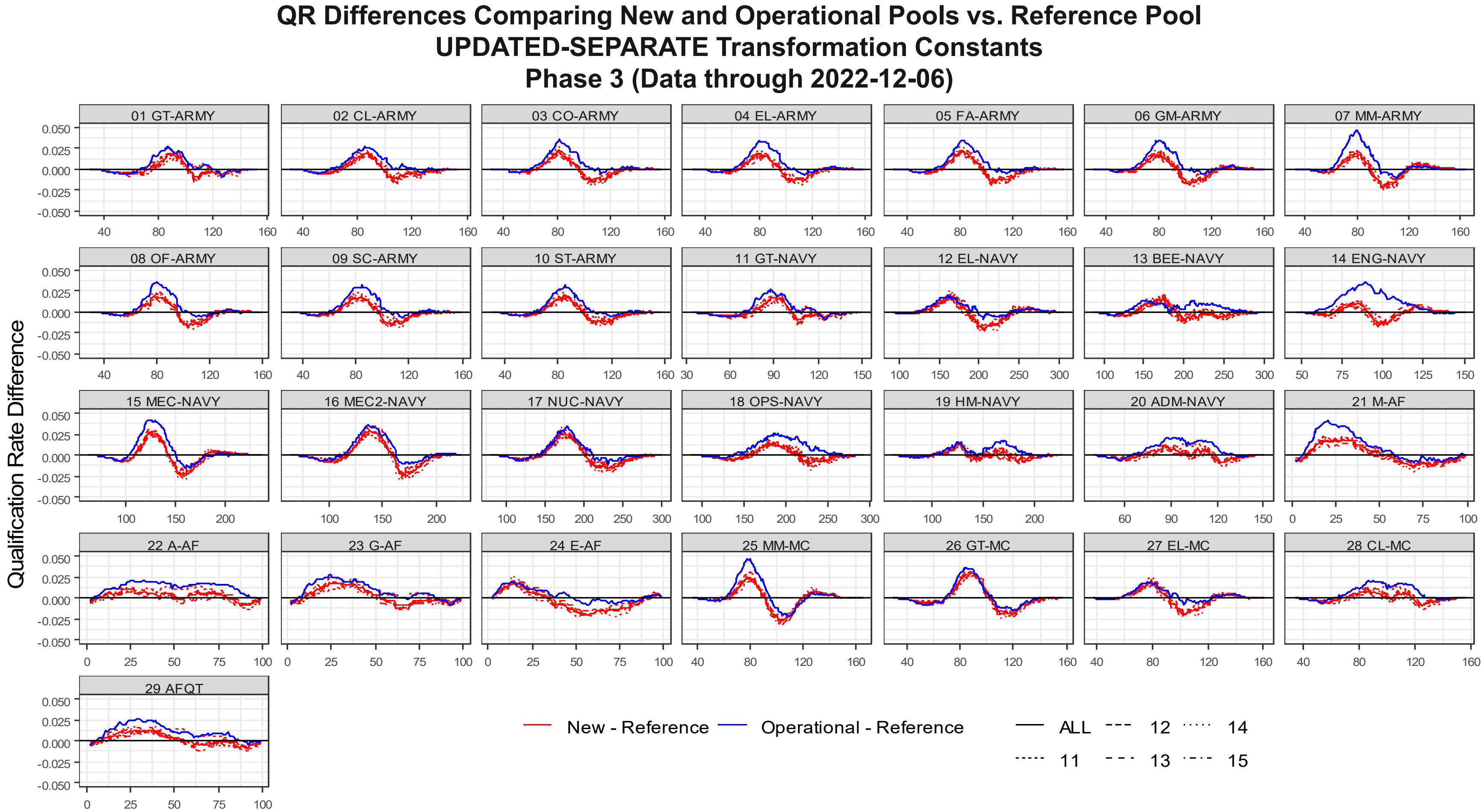
Composite Distribution Equivalence



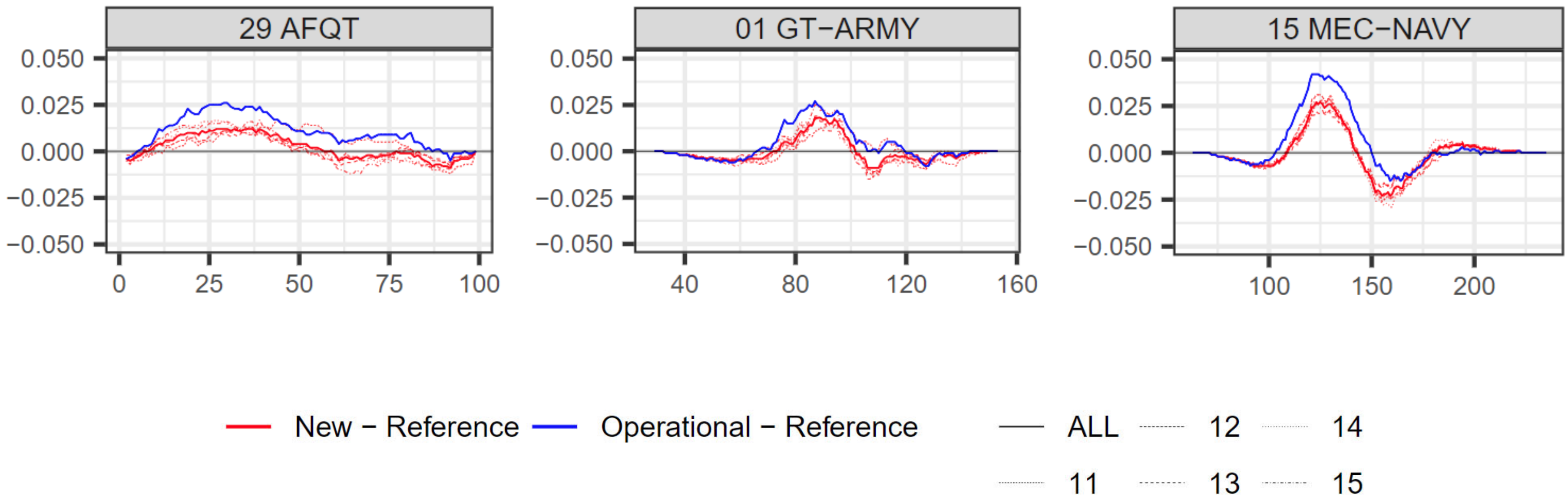
COMPOSITE DISTRIBUTION EQUIVALENCE



COMPOSITE DISTRIBUTION EQUIVALENCE



Composite Distribution Equivalence: Phase 3

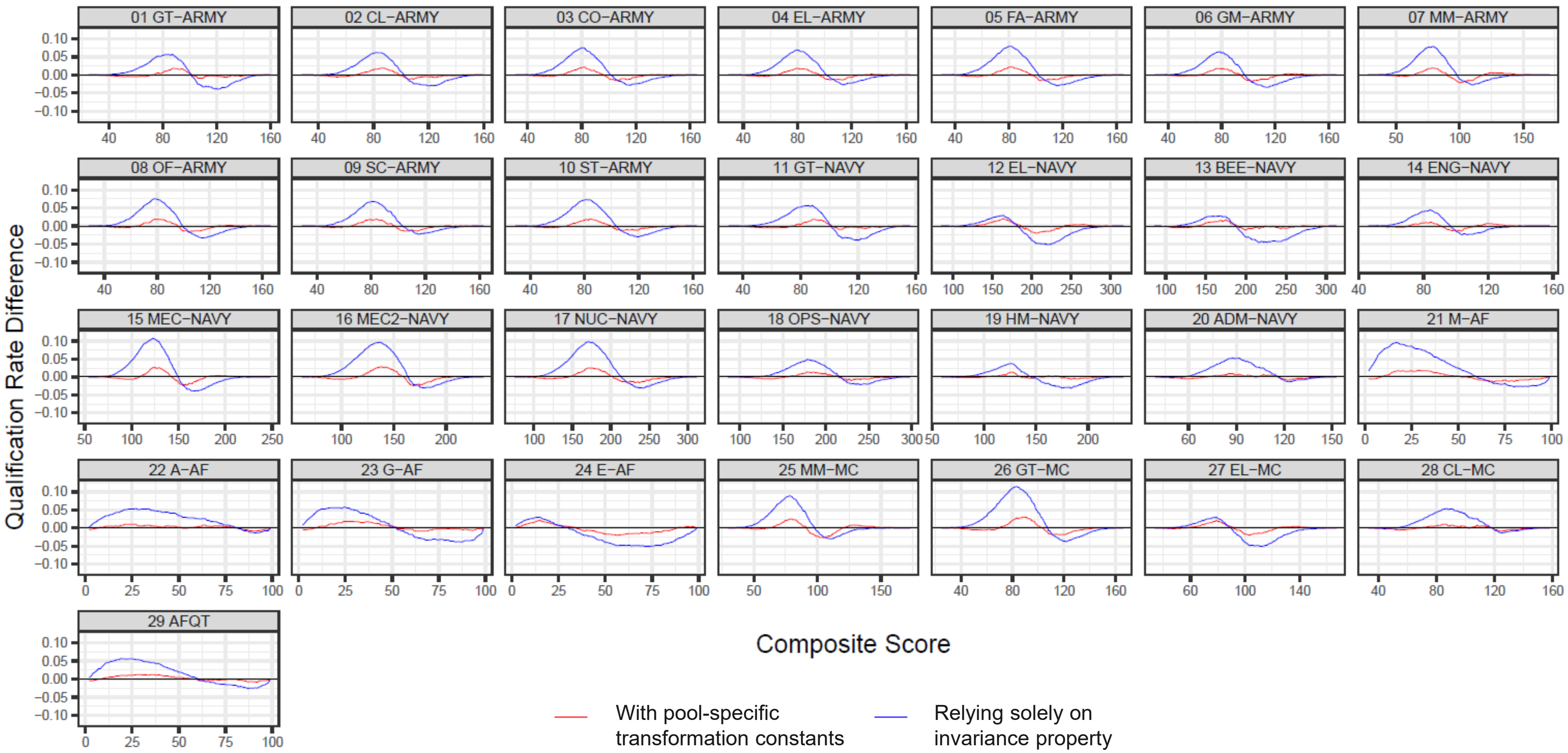


Equipercentile Objective and Invariance Revisited

- What if we relied solely on the IRT invariance property and did **not** estimate pool-specific transformation constants to match the mean and standard deviation of the reference pool?
  - See blue line in next slide (24)
  - In general, most composite distributions compare reasonably well to the reference pool
  - However, several composite distributions based on new pools deviate from the reference pool up to 10%, conditional on composite score
- Again, a large volume of applicants qualify on CAT-ASVAB, and small differences between (unequated) pools can potentially have a large impact on the number of qualified applicants
- This equating procedure ensures the equipercentile objective (i.e., distribution matching between a new pool standard scores and the reference pool standard scores), which enhances maximal similarity of qualification rates when a composite score is used for selection purposes
- The effectiveness of the equating procedure is evaluated for each new pool against the reference pool, and more importantly, in comparisons of score distributions between the new pool composite and the reference pool composite scores

EQUIPERCENTILE OBJECTIVE AND INVARIANCE REVISITED

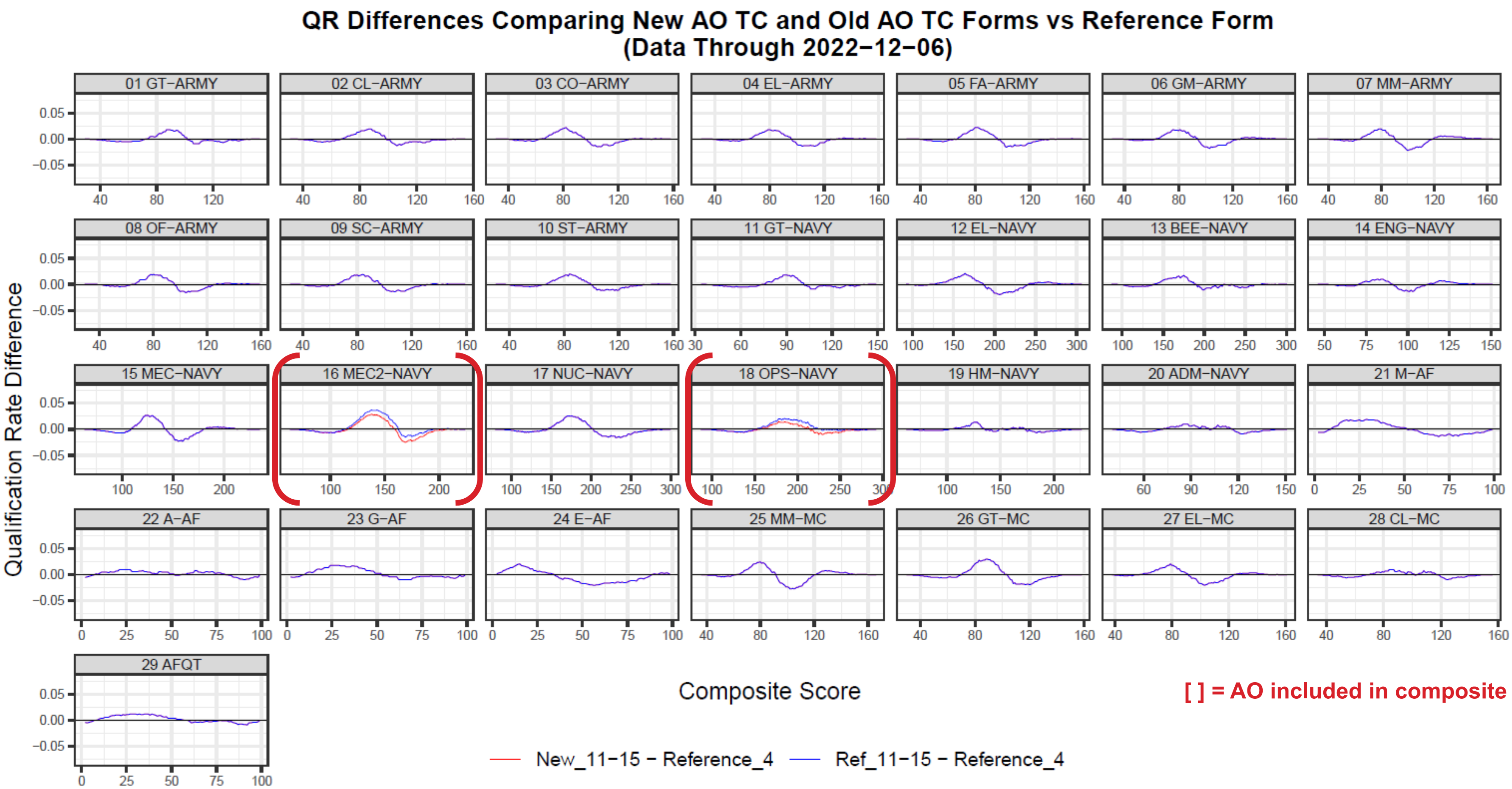
QR Differences Comparing New Transformation Constants & Reference Transformation Constants
Pools 11–15 vs. Reference Pool 4
(Data through 2022-12-06)



Assembling Objects (AO) Comparison

- Problems identified with Assembling Objects items considered for pools 11–15
  - Multidimensionality
  - Ceiling effects
- Re-used CAT-ASVAB pools 5–9 (relabeled as 11–15) while problems with next generation items addressed
  - Stand-alone MAPWG briefing on this topic (Waterbury, 2023)
- Two sets of standard score transformation constants
  - Original pools 5–9 equating study (estimated in 2008)
  - Current pools 11–15 equating study (estimated in 2022)
- AO is part of relatively few composites
  - MEC2-Navy, OPS-Navy included in evaluation template
- Original and new transformation constants produce very similar composite distributions
  - New constants are preferable for slightly improved similarity to reference pool

ASSEMBLING OBJECTS (AO) COMPARISON



Subgroup Performance

- Do subgroups perform at the same levels across pools?
  - Female examinees
  - Black examinees
  - Hispanic examinees
- Compare subgroup performance across new and reference pools
  - One-way ANOVA with groups defined by pool
  - Statistical significance and effect size

Subgroup Performance

Analysis of Pools 4 & 11–15:

- Female analysis: multiple statistically significant pairwise comparisons
  - AS: 3 differences (small effect sizes; $\delta = 0.13$, $\delta = 0.13$, and $\delta = 0.10$)
    - All statistically significant difference between pool 4 and new pools (11, 12, and 14, respectively)
    - No statistically significant differences between new pools
  - GS: 1 difference for pool 14 vs. 15 (small effect size; $\delta = 0.11$)
  - VE: 1 difference for pool 4 vs. 13 (small effect size; $\delta = 0.11$)
- Black examinee analysis: one statistically significant pairwise comparison
  - PC: 1 difference for pool 4 vs. 12 (small effect size; $\delta = 0.10$)
- Hispanic examinee analysis: no statistically significant pairwise comparisons
- Because all effect sizes are small, these results are not a concern
- These results are similar to those seen during pools 5–9 development
  - Similar number of statistically significant pairwise comparisons, and none that were practically significant

OPA OFFICE OF PEOPLE ANALYTICS

δ = Hedge's g

Operational Pool Comparison

How do equated scores on pools 11–15 compare to operational pool 5?

- Compared mean differences on operational **pool 5** to reference pool 4
- Statistically significant mean differences in several tests; however, all represented small effect sizes
- Pool 4 vs. 5
  - AFQT ($\delta = 0.04$)
  - AS ($\delta = 0.08$)
  - AO ($\delta = 0.07$)
  - MK ($\delta = 0.05$)
- Because all effect sizes are small, these results are not a concern
- These results are similar to those seen during pools 5–9 development
  - During pools 5–9 development, 3 subtests—including AO and MK—displayed statistically significant, but not practically significant, differences

OPA
OFFICE OF PEOPLE ANALYTICS

δ = Hedge's g

PROVISIONAL EQUATING TRANSFORMATION ACCURACY

- How closely did the provisional equating transformations match the final?
  - How different are scores based on provisional constants from what they would have been if based on final constants?
- Rescore all applicants who took pools 11–15, using final transformation constants
  - Compare rescored values to those used operationally based on provisional constants
- Calculate total errors as the sum of equating errors and measurement errors
- Compare total error with standard errors of measurement

PROVISIONAL TRANSFORMATION ACCURACY

Questions for the DAC

OPA
OFFICE OF PEOPLE ANALYTICS

QUESTIONS FOR THE DAC

- Does the DAC have feedback on potentially reducing the duration of phase 3 or even eliminating a phase?
- Other feedback on the equating or evaluation process?

HUMRRO PROJECT TEAM

- Maura Burke
- Jeff Dahlke
- Ted Diaz
- Olga Golovkina
- Ki Ho Kim
- Insu Paek
- Matthew Reeder
- Stephen Robertson
- Liz Waterbury

Technical Appendix

OFFICE OF PEOPLE ANALYTICS

Kolmogorov-Smirnov Test

* = significant difference according to K-S test ($p < .01$)

Service	Composite	04 vs 11	04 vs 12	04 vs 13	04 vs 14	04 vs 15	Average
AFQT	AFQT	0.017	0.012	0.013	0.017	0.016	0.015
Air Force	M	0.023*	0.019	0.016	0.022	0.022	0.020
	A	0.018	0.009	0.011	0.013	0.013	0.013
	G	0.017	0.020	0.018	0.019	0.025*	0.020
	E	0.021	0.020	0.021	0.026*	0.022	0.022
	Subtotal (of 4)	**1**	**0**	**0**	**1**	**1**	
Army	CL	0.017	0.020	0.017	0.025*	0.020	0.020
	CO	0.021	0.024*	0.018	0.026*	0.019	0.022
	EL	0.016	0.019	0.017	0.024*	0.019	0.019
	FA	0.021	0.023*	0.020	0.028*	0.022	0.023
	GM	0.016	0.021	0.018	0.022	0.019	0.019
	GT	0.017	0.020	0.018	0.019	0.025*	0.020
	MM	0.020	0.023*	0.023	0.026*	0.021	0.023
	OF	0.018	0.024*	0.018	0.024*	0.019	0.021
	SC	0.015	0.020	0.018	0.025*	0.020	0.020
	ST	0.016	0.021	0.017	0.026*	0.021	0.020
	Subtotal (of 10)	**0**	**4**	**0**	**8**	**1**	

Kolmogorov-Smirnov Test

* = significant difference according to K-S test ($p < .01$)

Service	Composite	04 vs 11	04 vs 12	04 vs 13	04 vs 14	04 vs 15	Average
Marine Corps	CL	0.018	0.009	0.011	0.013	0.013	0.013
	EL	0.021	0.020	0.021	0.026*	0.022	0.022
	GT	0.028*	0.031*	0.029*	0.033*	0.032*	0.031*
	MM	0.028*	0.031*	0.026*	0.031*	0.024*	0.028*
	Subtotal (of 4)	**2**	**2**	**2**	**3**	**2**	
Navy	ADM	0.018	0.009	0.011	0.013	0.013	0.013
	BEE	0.020	0.021	0.011	0.020	0.017	0.018
	EL	0.021	0.020	0.021	0.026*	0.022	0.022
	ENG	0.015	0.017	0.018	0.015	0.015	0.016
	GT	0.017	0.020	0.018	0.019	0.025*	0.020
	HM	0.014	0.014	0.011	0.016	0.016	0.014
	MEC	0.031*	0.031*	0.023*	0.030*	0.027*	0.028*
	MEC2	0.026*	0.032*	0.028*	0.030*	0.035*	0.030*
	NUC	0.023	0.027*	0.023*	0.033*	0.027*	0.027*
	OPS	0.013	0.015	0.016	0.016	0.018	0.016
	Subtotal (of 10)	**2**	**3**	**3**	**4**	**4**	
Total	**Total Differences**	**5**	**9**	**5**	**16**	**8**	**43**
Greatest Max. CDF Difference		**0.031**	**0.032**	**0.029**	**0.033**	**0.035**	

Provisional Equating Transformation Accuracy

RMSD, bias, and standard error of equating

Compute difference between Provisional and Final equating scores for each examinee $j = 1, \ldots, N$ and subtest $s = GS, AR, WK, PC, AI, SI, MK, MC, EI, AO, AS, VE$:

$$\delta_{s,j} = e_P(\hat{\theta}_{s,j}) - e_F(\hat{\theta}_{s,j})$$

For each subtest s:

$$\begin{aligned}
RMSD_s &= \left(\frac{1}{N}\sum_{j=1}^{N}\delta_{s,j}^2\right)^{1/2} \\
\bar{\delta}_s &= \frac{1}{N}\sum_{j=1}^{N}\delta_{s,j} \\
\sigma_{\delta_s} &= \left(\frac{1}{N}\sum_{j=1}^{N}\left(\delta_{s,j} - \bar{\delta}_s\right)^2\right)^{1/2}
\end{aligned}$$

OPA
OFFICE OF PEOPLE ANALYTICS

Provisional Equating Transformation Accuracy

Mean and SD of latent distribution of examinees

Find estimates $(\hat{\mu}, \hat{\sigma})$ of the mean and SD of examinee latent distribution that maximize the likelihood of observed responses given estimates of item parameters.

$$L(\mu, \sigma^2 | U) = \prod_{a=1}^{N} P\,(u_a | \mu, \sigma^2)$$

$$= \prod_{a=1}^{N} \int P(u_a | \theta) f(\theta | \mu, \sigma^2) d\theta$$

where $u_a = [u_{ai} | i = 1, \ldots, n]$ is a vector of item responses for examinee $a = 1, \ldots, N$ in the study, $f(\theta | \mu, \sigma^2)$ denotes a normal density, and

$$\prod_{a=1}^{N} \int P(u_a | \theta) = \prod_{i=1}^{n} [P(\theta | a_i, b_i, c_i)^{u_{ai}} Q(\theta | a_i, b_i, c_i)^{1-u_{ai}}]$$

for $i = 1, \ldots, n$ items with 3PL parameters a_i, b_i, c_i.

OPA
OFFICE OF PEOPLE ANALYTICS

Provisional Equating Transformation Accuracy

Average standard error of measurement (SEM) of ASVAB standard scores (SS)

- Compute average standard error of measurement of theta estimate

For subtest $s = GS, AR, WK, PC, AI, SI, MK, MC, EI, AO$

$$\sigma_{E,\theta,s} = \left[\frac{\sum_k w_k \, I(\theta_k, s)^{-1}}{\sum_k w_k}\right]^{1/2}$$

where $I(\theta_k, s)$ is the theta score information function of subtest s evaluated at θ_k and weights w_k are values of latent normal density $f(\theta|\hat{\mu}, \hat{\sigma}^2)$ evaluated at θ_k.

- Compute average SEM for ASVAB standard score:

Let $\sigma_{e_F,s}$ be the standard deviation of the final updated standard score and $\hat{\sigma}_s$ be the SD of the latent distribution for subtest s. For subtest $s = GS, AR, WK, PC, AI, SI, MK, MC, EI, AO$ the average SEM is obtained by

$$\sigma_{E,X,s} = \frac{\sigma_{E,\theta,s}}{\hat{\sigma}_s} \sigma_{e_F,s}$$

For AS and VE standard scores the average SEM is obtained from component test below.

$$\sigma_{E,X,VE} = \frac{1}{2}\sqrt{\sigma^2_{E,X,WK} + \sigma^2_{E,X,PC}}$$

$$\sigma_{E,X,AS} = \frac{1}{2}\sqrt{\sigma^2_{E,X,AI} + \sigma^2_{E,X,SI}}$$

OPA
OFFICE OF PEOPLE ANALYTICS

Provisional Equating Transformation Accuracy

Total error and incremental error due to equating

Compute total error as the sum of equating error and measurement error:

$$\sigma_{E_{Tot},X,s} = \sqrt{\sigma^2_{\delta_s} + \sigma^2_{E,X,s}}$$

Compute absolute and relative incremental error due to equating

$$\sigma_{E_{Inc},X,s} = \sigma_{E_{Tot},X,s} - \sigma_{E,X,s}$$

$$\sigma_{E_{PInc},X,s} = \frac{\sigma_{E_{Inc},X,s}}{\sigma_{E_{Tot},X,s}}$$

OPA
OFFICE OF PEOPLE ANALYTICS

Provisional Transformation Accuracy

Accuracy of IRT Equating (Used in Phase I) Averaged Across Pools							
Test	**RMSD**	**Bias**	σ_{δ_s}	$\sigma_{E,X,s}$	$\sigma_{E_{Tot},X,s}$	$\sigma_{E_{Inc},X,s}$	$\sigma_{E_{PInc},X,s}$
GS	5.13	1.59	1.57	3.80	4.11	0.31	7.59
AR	4.95	0.68	2.11	3.51	4.09	0.58	14.25
WK	0.54	-0.42	0.57	3.03	3.09	0.06	1.79
PC	10.14	-2.38	2.10	4.28	4.77	0.49	10.27
MK	0.49	-0.20	0.61	3.47	3.53	0.06	1.66
MC	16.78	-3.15	2.61	4.69	5.37	0.68	12.61
EI	0.69	-0.31	0.74	5.06	5.11	0.06	1.09
AS	2.67	-0.22	1.59	4.98	5.23	0.25	4.89
VE	2.84	-1.23	1.13	2.64	2.88	0.24	8.14

Provisional Transformation Accuracy

Accuracy of Equating Developed in Phase I (Used in Phase II) Averaged Across Pools							
Test	**RMSD**	**Bias**	σ_{δ_s}	$\sigma_{E,X,s}$	$\sigma_{E_{Tot},X,s}$	$\sigma_{E_{Inc},X,s}$	$\sigma_{E_{PInc},X,s}$
GS	0.47	0.41	0.47	3.82	3.85	0.03	0.76
AR	0.47	0.34	0.56	3.46	3.51	0.05	1.30
WK	0.47	0.31	0.58	3.02	3.07	0.06	1.83
PC	0.40	0.39	0.43	4.17	4.19	0.02	0.54
MK	0.25	0.02	0.44	3.45	3.48	0.03	0.81
MC	0.40	0.38	0.50	4.68	4.70	0.03	0.57
EI	0.66	0.35	0.69	5.13	5.18	0.05	0.92
AS	1.02	0.90	0.41	4.95	4.97	0.02	0.35
VE	0.41	0.36	0.51	2.61	2.66	0.05	1.89

Provisional Transformation Accuracy

Accuracy of Equating Developed in Phase II (Used in Phase III) Averaged Across Pools							
Test	**RMSD**	**Bias**	σ_{δ_s}	$\sigma_{E,X,s}$	$\sigma_{E_{Tot},X,s}$	$\sigma_{E_{Inc},X,s}$	$\sigma_{E_{PInc},X,s}$
GS	0.32	0.24	0.42	3.81	3.84	0.03	0.65
AR	0.20	-0.08	0.40	3.45	3.47	0.02	0.67
WK	0.34	0.02	0.55	3.03	3.08	0.05	1.69
PC	0.24	0.04	0.41	4.23	4.25	0.02	0.48
MK	0.33	0.03	0.53	3.45	3.49	0.04	1.16
MC	0.12	0.01	0.31	4.68	4.69	0.01	0.25
EI	0.41	0.34	0.52	5.09	5.11	0.03	0.52
AS	0.27	0.21	0.43	5.04	5.05	0.02	0.37
VE	0.25	0.03	0.47	2.64	2.68	0.04	1.59

Provisional Transformation Accuracy

Accuracy of Provisional Equating (Used Across Phases I–III) Averaged Across Pools							
Test	RMSD	Bias	σ_{δ_s}	$\sigma_{E,X,s}$	$\sigma_{E_{Tot},X,s}$	$\sigma_{E_{Inc},X,s}$	$\sigma_{E_{PInc},X,s}$
GS	0.81	0.39	0.76	3.81	3.89	0.08	1.93
AR	0.69	0.04	0.81	3.46	3.55	0.09	2.63
WK	0.38	0.01	0.58	3.03	3.08	0.06	1.86
PC	1.22	-0.16	1.06	4.23	4.36	0.13	3.02
MK	0.34	0.01	0.54	3.45	3.49	0.04	1.20
MC	1.77	-0.26	1.30	4.68	4.86	0.18	3.68
EI	0.47	0.28	0.60	5.09	5.12	0.04	0.69
AS	0.58	0.24	0.70	5.02	5.07	0.05	0.96
VE	0.52	-0.05	0.70	2.63	2.73	0.09	3.38

OPA
OFFICE OF PEOPLE ANALYTICS

Thank you!

For more information please contact:

Jeff Dahlke
jdahlke@humrro.org

OFFICE OF PEOPLE ANALYTICS